

**W35 – 1st Multimodal Learning
and Applications Workshop**

W35 – 1st Multimodal Learning and Applications Workshop

The exploitation of the power of big data in the last few years led to a big step forward in many applications of Computer Vision. However, most of the tasks tackled so far are involving mainly visual modality due to the unbalanced number of labeled samples available among modalities (e.g., there are many huge labeled datasets for images while not as many for audio or IMU based classification), resulting in a huge gap in performance when algorithms are trained separately. The workshop addressed various themes in the field of multimodal data fusion and representation (e.g. we had contribution proposing method to deal with modality fusion between Audio-Video, Video-Lidar, Text-Image, etc.). As expected we have attracted researcher from different environments, we had nearly equal contributions from academia and industry.

As part of the program we had the privilege to have two eminent invited speakers: Daniel Cremers from TU Munich and Raquel Urtasun from Uber ATG Toronto. In their talks they have shown how multimodal learning is a topic of tremendous importance, for industrial applications such as automotive and SLAM technologies. The workshop, despite its first edition, has been very successful. We received 28 valid submissions all reviewed by at least two reviewers resulting in 11 papers accepted (39% acceptance rate). Three paper have been presented as oral while the other 8 as poster plus a brief spotlight session. Given the quality of the papers we have decided to propose a best paper award which has been sponsored by Bosch.

We would like to express our gratitude to all the Authors, to the Invited Speakers and to the members of the Program Committee for the amazing contribution they have done to the workshop.

September 2018

Paolo Rota
Vittorio Murino
Michael Yang
Bodo Rosenhahn